Non-adaptive origins of evolutionary innovations increase network complexity in interacting digital organisms

Fortuna, Miguel A ; Zaman, Luis ; Wagner, Andreas ; Bascompte, Jordi

Abstract: The origin of evolutionary innovations is a central problem in evolutionary biology. To what extent such innovations have adaptive or non-adaptive origins is hard to assess in real organisms. This limitation, however, can be overcome using digital organisms, i.e. self-replicating computer programs that mutate, evolve and coevolve within a user-defined computational environment. Here, we quantify the role of the non-adaptive origins of host resistance traits in determining the evolution of ecological interactions among host and parasite digital organisms.We find that host resistance traits arising spontaneously as exaptations increase the complexity of antagonistic host–parasite networks. Specifically, they lead to higher host phenotypic diversification, a larger number of ecological interactions and higher heterogeneity in interaction strengths. Given the potential of network architecture to affect network dynamics, such exaptationsmay increase the persistence of entire communities. Our in silico approach, therefore, may complement current theoretical advances aimed at disentangling the ecological and evolutionary mechanisms shaping species interaction networks. This article is part of the themed issue 'Process and pattern in innovations from cells to societies'.

DOI: https://doi.org/10.1098/rstb.2016.0431

Non-adaptive origins of evolutionary innovations increase network complexity in interacting digital organisms*

Miguel A. Fortuna[1]†, Luis Zaman[2,3], Andreas Wagner[1,4,5], and Jordi Bascompte[1‡]

[1]Department of Evolutionary Biology and Environmental Studies
University of Zurich, Zurich (Switzerland)

[2]Department of Biology,
University of Washington, Seattle, Washington (USA)

[3]BEACON Center for the Study of Evolution in Action,
Michigan State University, East Lansing, Michigan (USA)

[4]Swiss Institute of Bioinformatics
Lausanne (Switzerland)

[5]The Santa Fe Institute
Santa Fe, New Mexico (USA)

*Invited chapter for the special issue on *Process and Pattern in Innovations from Cells to Societies* of *Philosophical Transactions B*.

†Corresponding author: miguel.fortuna@ieu.uzh.ch

‡Corresponding author: jordi.bascompte@ieu.uzh.ch

Abstract

The origin of evolutionary innovations is a central problem in evolutionary biology. To what extent such innovations have adaptive or non-adaptive origins is hard to assess in real organisms. This limitation, however, can be overcome using digital organisms, i.e., self-replicating computer programs that mutate, evolve, and coevolve within a user-defined computational environment. Here we quantify the role of the non-adaptive origins of host resistance traits in determining the evolution of ecological interactions among host and parasite digital organisms. We find that host resistance traits arising spontaneously as exaptations increase the complexity of antagonistic host-parasite networks. Specifically, they lead to higher host phenotypic diversification, larger number of ecological interactions, and higher heterogeneity in interaction strengths. Given the potential of network architecture to affect network dynamics, such exaptations may increase the persistence of entire communities. Our "in silico" approach, therefore, may complement current theoretical advances aimed at disentangling the ecological and evolutionary mechanisms shaping species interaction networks.

Keywords: digital coevolution, ecological networks, host-parasite interactions, exaptation.

Introduction

It has recently been shown that interactions among coevolving species promote the emergence of evolutionary innovations, defined as qualitatively novel and beneficial traits. Among such innovations are host resistance traits for escaping parasites (Zaman *et al.*, 2014) and the ability of parasites to infect either new (Duffy *et al.*, 2007; Elena, 2016) or current (Meyer *et al.*, 2012) hosts through novel pathways. Some of these studies have even identified the sequence of mutations leading to evolutionary innovation (e.g., Duffy *et al.*, 2007; Meyer *et al.*, 2012; Elena, 2016). In general, the larger the number of mutations required to evolve an innovation, the less likely is that this process takes place in a single step (e.g., Kuinen *et al.*, 2006). However, we still know little about the evolutionary origins—adaptive or non-adaptive—of such innovations, and whether they foster the role of coevolution in opening multiple paths to future innovation.

Palaeontologists Stephen J. Gould and Elisabeth S. Vrba (1982) introduced the concept of exaptations to refer to organismal traits that are either non-adaptive when they originate, or that were selected for a different function than the one currently performed. For example, the evolution of the genome complexity from prokaryotes to multicellular eukaryotes might have non-adaptive origins (Lynch & Conery, 2003). This non-adaptationist theory is supported by the increase in genome entropy, which is inevitably triggered by reduction of population size—which in turn, strengthened the effects of random genetic drift and weaken the effects of purifying selection (Koonin, 2004). In fact, it has been shown that small and large populations are favored to evolved larger genomes, which provides the opportunity for subsequent increases in phenotypic complexity (Labar & Adami, 2016). More recently, experimental studies on promiscuous enzymes (O'Brien & Herschlag, 1999; Aharoni *et al.*, 2005) have emphasized the importance of exaptation in

evolution. These proteins can acquire new functions as byproducts of adaptations and thus help organisms survive in different environments. Similarly, recent work on metabolic networks has focused on how often adaptive metabolic traits have non-adaptive origins (Barve & Wagner, 2013; Notebaart *et al.*,, 2014; Hosseini & Wagner, 2016). These studies showed that bacteria viable on glucose as a sole carbon source can also be viable on multiple other carbon sources that were not targets of selection. This non-adaptive ability of surviving in alternative carbon sources emerges as a byproduct of the complexity of biochemical reaction networks. Indeed, the complexity of metabolic networks can increase the potential for exaptations and, hence, can thus contribute to the pervasiveness of non-adaptive traits in biological systems (Barve & Wagner, 2013).

In artificial life systems, such as self-replicating and evolving computer programs—digital organisms—the ability of an organism to compute simple Boolean logic functions on binary numbers can emerge, likewise, as a byproduct of computing more complex functions. These genetically-encoded phenotypes result from the coordinated execution of "genetic building blocks" (i.e., computational instructions that organisms harbor in their genomes), which are analogous to developmental processes guided by regulatory programs in biology (Fortuna *et al.*, 2017). The higher the complexity of a function computed by a digital organism is, the greater is the likelihood that the organism can also compute simpler functions, i.e., the greater is the potential for exaptations.

If qualitatively novel and potentially beneficial traits arise spontaneously and non-adaptatively, regardless of any later adaptive function, the environment may determine how fast they become adaptive. On the one hand, biotic interactions might play an important role in determining the benefit that a trait provides and help natural selection spread it through a population. Recent coevolutionary models have suggested that species interactions in complex networks change the mean value of the traits involved in the eco-

logical interactions among the partners (Guimarães *et al.*, 2011; Nuismer *et al.*, 2013; Andreazzi *et al.*, 2017), which influences evolutionary dynamics. On the other hand, a frequent non-adaptive origin of evolutionary innovations might drive species interactions and enhance the complexity of the entangled web of ecological interactions among organisms. Quantifying the role of exaptations in shaping species interaction networks requires a framework to discern exaptations from adaptations.

Disentangling non-adaptive from adaptive origins of evolutionary innovations in natural ecological communities is so far unfeasible. In contrast, artificial life evolving systems—such as the digital organisms mentioned above—allow us to suppress mutations responsible for non-adaptive origins of evolutionary innovations. Avida is a widely-used software platform for the study of evolution (see Wilke & Ofria, 2004) that has recently been extended to study host-parasite coevolution (Fortuna *et al.*, 2013; Zaman *et al.*, 2014).

Digital coevolution between hosts and parasites resembles the coevolutionary dynamics among *Escherichia coli* and lambda phages (Fig. 1). On the one hand, bacteria must have receptors on their surface in order to import resources from their environment. On the other hand, phages must attach to those receptors in order to infect bacteria. Therefore, a trade-off exists between having receptors for obtaining nutrients and being susceptible to phages. Coevolutionary dynamics results from bacteria evolving phage resistance by changing their surface receptors, and phages countering resistance by altering their tail fibers to attach to the novel receptors (Meyer *et al.*, 2012). Analogously, digital hosts must compute logic operations to consume resources and thus replicate, but those traits leave them susceptible to infection by digital parasites. Here we use digital coevolution to shed light on the role of the non-adaptive origins of evolutionary innovations in shaping the web of life.

Methods

Digital evolution

Digital evolution is an applied branch of Artificial Life. In this evolutionary computation framework, self-replicating computer programs—digital organisms—evolve within a user-defined computational environment (Wilke & Adami, 2002). Avida is the most widely used software platform for research in digital evolution (Ofria & Wilke, 2004). It satisfies the three essential requirements for evolution to occur: replication, heritable variation, and differential fitness. Differences in fitness among digital organisms arise through competition for the limited resources of memory space and central processing unit (CPU) time. A digital organism in Avida consists of a sequence of instructions—its genome or genotype—and a virtual CPU, which executes these instructions. Some of these instructions are involved in copying an organism's genome, which is the only way the organism can pass on its genetic material to future generations. To reproduce, a digital organism must copy its genome instruction by instruction into a new region of memory. The copying process occasionally introduces mutations including point mutations, insertions, and deletions. For example, a point mutation occurs when an instruction is copied incorrectly, and is instead replaced in the offspring genome by an instruction chosen at random (with a uniform distribution) from a set of 33 possible instructions. In addition to the instructions required for replication (i.e., viability), the instruction set includes basic arithmetic operations (such as addition, multiplications, and bit-shifts) as well as the logic operator *nand* that are executed on binary numbers taken from the environment through input-output instructions. When the output of processing these numbers equals the result of a specific Boolean logic operation, such as the AND and OR Boolean functions, the digital organism is said to have a trait represented by that logic operation. We have focused

here on the following 9 Boolean logic operations that organisms can perform on 32-bit one- and two-input numbers: NOT, NAND (not-and), AND, OR_N (or-not), OR, AND_N (and-not), NOR (not-or), XOR (exclusive or), and EQU (logical equality).

We configured the environment so that it contains a single resource that must be consumed by organisms for successful replication. This resource renews at a constant rate and influences the carrying capacity of the population. For organisms to successfully consume a unit of resource, they must compute before replication, at least one Boolean logic function. If there is no single available unit of this resource, replication fails and the organism begins the execution of its genome's instructions again without producing any offspring. When applied to multiple organisms, this procedure leads to density-dependent population growth. In addition, there are two sources of density-dependent mortality: 1) if an organism does not successfully divide, it dies after executing 5 times its genome's instructions (i.e., 5 failed replication attemps); and 2) as a result of offspring being randomly placed in the population, a resident organism can be overwritten by the newly generated offspring.

Digital coevolution

In order to study coevolution in populations of interacting organisms, we have implemented parasitic organisms and a mechanism for them to infect the above described digital organisms (i.e., hosts) based on genetically encoded phenotypes (Zaman *et al.*, 2011; 2014). This new branch of Avida supports threading capabilities (i.e., more than one type of organism executing their instructions in parallel) and separate memory spaces for hosts and parasites (i.e., regions of memory reserved for the genome of the offspring produced during the self-replication process). It uses a 33-instruction set that expanded the 26-instruction default genetic language of Avida (Supplementary File 1).

Parasitic digital organisms are almost identical to the hosts, and as such they self-replicate by copying their genome instruction-by-instruction into a new memory space. But they operate inside hosts, stealing CPU cycles from them to execute their own genome's instructions and, hence, reduce their host fitness. However, parasites have memory spaces entirely separated from their host's and, therefore, do not have access to their host's instructions. This avoids any unforeseen side-effects such as those observed in the digital evolution platform Tierra (Ray, 1991), where parasites can overwrite their host's genome. The selective pressure of a parasite on its host (i.e., virulence) is determined by the probability that a CPU cycle from the host will be given to the parasite. When this probability is set to 0.5, parasites and hosts split CPU cycles evenly, and when it is set to 1, the parasite uses all of its host's CPU cycles. In the latter case, the relationship is analogous to a predator-prey interaction. After making a copy of its genome using the CPU cycles "stolen" from its host, a parasite must place its newly generated offspring into an uninfected host. It attemps to do this by executing the instruction "Inject". The parasite's offspring is then randomly placed in the host population. If the chosen memory space is occupied by an uninfected host, the parasite can infect it. In contrast, if the memory space is either occupied by an infected host or empty, infection will fail. Multiple infections are not allowed (i.e., a host can only be infected by one parasite).

Parasites can not infect just any host. The mechanism of infection is based on phenotypic trait matching, i.e., an uninfected host will be infected by a parasite's offspring if the parasite computed at least one of the Boolean logic functions that the host also computed. This mechanism of infection emulates the inverse gene-for-gene model, where infectiousness is determined by parasite recognition of host signals and/or receptors (Fenton *et al.*, 2009). Parasites, like hosts, show density-dependent population growth (i.e., infections fail when most hosts are already infected). In addition, a parasite dies when

its host replicates successfully.

Coevolutionary dynamics

We sampled genotype space to find the hosts and parasites that were used as ancestors in the coevolutionary scenarios described below. Specifically, we first identified 30 mutationally-robust viable hosts capable of performing only the NOT logic operation and 15 mutationally-robust viable parasites capable of performing only the NOT logic operation. Next, we quantified the long-term stable coexistence of all $30 \times 15 = 450$ host-parasite pairs in a purely ecological scenario where mutations were allowed neither in hosts, nor in parasites. We kept the 216 (48%) host-parasite pairs where both hosts and parasites coexisted to study their coevolutionary dynamics (see Supporting Information).

Then, we expanded this purely ecological framework by introducing evolution into host-parasite population dynamics. We introduced novel genotypic variation into the host and parasite populations as single-point mutations, i.e., substitutions of one instruction in the offspring's genome by another instruction randomly chosen from the 33-instruction set. Hence, genome size was kept constant for both hosts and parasites. We applied a mutation rate per instruction in an organism's genome of $\mu_H = 0.025$ and $\mu_P = 0.01$ for hosts and parasites, respectively. This means that, on average, 1 out of 40 host offpring will become a novel host, and 1 out of 100 parasite offspring will become a novel parasite. Note that in nature, mutation rates of phages are higher than those of bacteria but their genomes are also smaller, while here, genome size was the same and kept constant during the evolutionary processes for both hosts and parasites. We set the probability that a CPU cycle from an infected host was given to the parasite to 0.9 (i.e., the parasite used 90% of the host CPU cycles).

For each of the 216 host-parasite pairs that coexisted in the long term in a purely

ecological scenario (Supplementary File 2), we performed 10 independent coevolutionary processes (i.e., replicates). Note that both host and parasite ancestors were able to compute the Boolean logic function NOT. Each process started from a population of 10^4 hosts with the same ancestral genotype and, after the host reached its carrying capacity ($K \approx 6500$, on average for the 30 distinct hosts used), we infected half of the host population with the same ancestral parasite genotype. After 2×10^5 updates, where an update is the amount of time during which an organism executes on average 30 instructions (i.e., on average 10^4 generations), we stopped the coevolutionary process and retrieved the data generated during the entire process. We repeated each coevolutionary process for each of the two scenarios described below, i.e., allowing and prohibiting non-adaptive origins of resistance traits ($216 \times 10 \times 2 = 4320$ coevolutionary processes).

Non-adaptive and adaptive origins of resistance traits

As a host population evolves, organisms might perform other Boolean logic operations besides or instead of NOT. Performing a new logic function is analogous to evolving a novel trait. In order to become resistant, hosts must lose their ability to perform the ancestral NOT logic function (so that parasites cannot infect them) while also evolving a novel trait (i.e., the ability to perform a new Boolean logic function) so that they can continue to collect the resources required for replication. If a host evolves a novel logic function without losing the ancestral one, the novel function has no adaptive value, since the parasite can still infect the host, and since there is no fitness advantage in having more the one trait for collecting resources. Later on, when a host loses the ability to perform the ancestral logic function, the evolved trait will have an adaptive value since it has become necessary to collect resources. We then refer to this evolved trait as an exaptation—an adaptive trait of an organism that was not adaptive when it originatedor we say that

this adaptation has a non-adaptive origin. In contrast to the scenario described next, we allow the replication of hosts capable of performing a novel logic function while keeping the ability to perform the ancestral one (Supplementary File 3).

When a host evolves a novel logic function while losing its ability to perform the ancestral one, the novel trait has adaptive value from its inception. This is so because it allows the host to collect resources required for replication and confers resistance to parasites. We supress exaptations in this scenario by preventing the replication of hosts that perform more than one logic function (Supplementary File 4). Novel traits evolve at the same time that hosts lose ancestral traits. This means that a single-point mutation is reponsible for both acquiring the novel function and loosing the ancestral one. By preventing the replication of hosts having more than one trait we might bias the evolutionary trajectories that the host population could follow. However, this represents only 3.67% of the population (i.e., $\approx$ 239 organisms in a population of $\approx$ 6500 hosts). Moreover, the fraction of the 1-mutant neighbors that are non-viable (i.e., they have genomes that do not allow them to produce offsprings) is smaller when traits have also non-adaptive origins (28%) than when they have only adaptive origins (31%; see Supplementary Figure 1). This reduction in host population abundance introduced by new mutations is 10 orders of magnitude larger than the 3.67% decrease in host population size artificially induced by our experimental design when traits have only adaptive origins (i.e., preventing the replication of hosts having more than one trait). Therefore, the slowing down in the host evolutionary potential when traits have only adaptive origins comes naturally from the coevolutionary process and not as much from our experimental procedure to prevent exaptations.

Results

Evolutionary innovations that emerged as exaptations—adaptive traits that were not adaptive when they originated—influenced the coevolutionary dynamics among host and parasite digital organisms (see Fig. 2). Specifically, they altered the following ecological and evolutionary responses: 1) the likelihood for hosts to escape from parasites; 2) population abundances; 3) the evolution of host resistance traits; and 4) the complexity of the network of interactions among hosts and parasites.

Exaptations and the likelihood of hosts escaping parasites

We observed coexistence of hosts and parasites during the entire coevolutionary process in both scenarios, namely, when resistance traits had both adaptive and non-adaptive origins vs when they had only adaptive origins. Specifically, for 90% of the 216 host-parasite pairs used as ancestors of the coevolutionary processes, host and parasite populations survived in the long-term. Parasites did not drive host extinction during any coevolutionary process. In contrast, hosts escaped from parasites—driving them to extinction—more frequently when resistance traits had both adaptive and non-adaptive origins than when they had only adaptive origins (it happened in at least one coevolutionary process for 84% and 76% of the 216 host-parasite pairs used as ancestors, respectively). That is, when resistance traits have non-adaptive origins, it is more likely for hosts to escape from parasites ($\chi^2 = 3.72$, $df = 1$, $p = 0.027$; two-sample test for equal proportions).

Exaptations and population abundance

Interestingly, the fraction of infected hosts—averaged over the entire coevolutionary process and over replicates where hosts and parasites coexisted—was 4% higher when the traits had a non-adaptive origin than when they had to have adaptive value from their in-

ception ($t = 5.51$, $df = 192$, $p < 0.001$; paired t-test). This result contrasts with observed patterns in population abundance. Indeed, host and parasite population abundances—averaged in the same way—were 15% and 18% higher, respectively, when resistance traits had non-adaptive origins ($t = 15.88$, $df = 192$, $p < 0.001$, and $t = 17.87$, $df = 192$, $p < 0.001$, respectively; paired t-test; Fig. 3a-b).

Exaptations and the evolution of host resistance traits

The maximum number of Boolean logic functions (i.e, potentially beneficial traits) evolved by hosts across all coevolutionary processes where they coexisted with the parasites was 7 (out of 9, which is the maximum number of traits), regardless of the adaptive origin of those traits. However, when we compared the maximum number of novel traits evolved by hosts initiated from the same host-parasite pair in both scenarios, we found a highly significant 23% increase in the number of evolved traits with non-adaptive origin ($t = 7.05$, $df = 192$, $p < 0.001$; paired t-test). That is, the non-adaptive origin of resistance traits promotes host phenotypic diversification (Fig. 3c).

Exaptations and the complexity of host-parasite networks

The number of interactions between single-trait host phenotypes and parasite phenotypes (i.e., pairwise infections)—averaged over the entire coevolutionary process and over replicate where hosts and parasites coexisted—was twice as high when the traits had also non-adaptive origins ($t = 10.91$, $df = 192$, $p < 0.001$; Fig. 3d). Beyond the number of ecological interactions, we calculated interaction evenness (Bersier *et al.*, 2002). It measures how equifrequently distributed interaction strengths are among hosts and parasites (accounting for differences in the number of interactions). Interaction strengths are estimated as the fraction of hosts with a given phenotype that is infected by each parasite

phenotype. The opposite of this measure gives an idea of the heterogeneity in the distribution of interaction strengths across links. Such heterogeneity in interaction strengths was higher (i.e., interaction evenness was 6% lower) when traits were allowed to have non-adaptive origins than when they have only adaptive origins ($t = -2.97$, $df = 192$, $p = 0.002$; Fig. 3e).

Discussion

We have shown that the non-adaptive origins of host resistance traits facilitate hosts escaping from parasites and promote host phenotypic diversification. This ability of hosts to acquire qualitatively novel and beneficial functions (i.e., evolutionary innovations) enables parasites to expand their ecological opportunities. As a result, the role of evolutionary innovations extends beyond the pairwise realm to entire networks of interactions. Specifically, we have shown that exaptations promote the complexity of antagonistic host-parasite networks by increasing the number of interactions among host and parasite phenotypes and the heterogeneity across interaction strengths.

In a previous study we have shown that potential exaptations appear spontaneously in evolving digital organisms and help bridge mutationally-connected networks of genotypes having the same phenotype (i.e., genotype networks; see Fortuna *et al.*, 2017). This high prevalence of exaptations suggests that the likelihood for a population to reach a novel trait j from organisms having trait i ($p_{i \to j}$) might be greater if the population encounters trait j first as an exaptation (i.e., organisms can have both traits at the same time) than finding directly the novel trait j (i.e., $p_{i \to ij} \times p_{ij \to j} > p_{i \to j}$). This could explain why the non-adaptive origins of evolutionary innovations facilitates hosts escape from parasites.

The presence of exaptations increases population abundances for both hosts and parasites, which can make the entire host-parasite network more robust to stochastic

fluctuations. A potential mechanism that might explain why the abundance of the host population is larger when resistance traits have adaptive and non-adaptive origins than when they have only adaptive origins relies on the characteristics of the regions of the genotype space occupied by the evolving host populations in both scenarios. When exaptations are allowed, the host population move towards a more robust region of genotype space (i.e., the fraction of the 1-mutant neighbors having the same phenotype as the target organism is larger when traits have also non-adaptive origins than when they have only adaptive origins; see Supplementary Figure 1).

Exaptations seem also to affect network robustness through their effect on our two measures of network architecture. First, the additional number of links resulting from the non-adaptive origins of innovations results in more connected networks. Such networks generally are more robust to species extinctions (Dunne *et al.*, 2002). Second, the increase in the heterogeneity of the distribution of interaction strengths resembles patterns in food webs and mutualistic networks, which contain a few strong interactions embedded within a matrix of weak interactions (Paine, 1992; Fagan & Hurd, 1994; Rafaelli & Hall, 1995; Wootton, 1997; Bascompte *et al.*, 2005; 2006). This heterogeneity has also been found to increase network stability (McCann *et al.*, 1998; Kokkoris *et al.*, 1999; Neutel *et al.*, 2002; Bascompte *et al.*, 2005; 2006).

One can perceive the patterns in network architecture reported above as an evolutionary innovation that increases network persistence. Thus, the direct evolutionary innovations at the scale of novel species interactions mediated through exaptations can scale up to generate a novel type of innovation, that of network structures that mediates the persistence of entire networks. Interaction networks shaped by exaptations, thus, should be view as coevolved structures that can not be reduced to the simple addition of the constituent pairwise interactions.

Network research has emphasised that understanding the relationships between pairs of species is not enough to understand the coexistence and functioning of entire networks. Our results show that coevolutionary processes may increase the number of opportunities leading to innovations of non-adaptive origin, which generates more complex networks. Coevolution, therefore, seems to be relevant for understanding the mechanisms shaping the complex web of life (Guimarães *et al.*, 2011; Nuismer *et al.*, 2013). Our results on the role of exaptations for ecological communities are also in line with Prigogine's ideas of self-organization, by which innovations occur stochastically and are integrated into a system by deterministic relationships existing at the time (Prigogine, 1980).

Acknowledgements

We thank the Science Cloud computing infrastructure of the University of Zurich for computational resourcess, and Matt Barbour for helpful comments on a previous draft. AW acknowledges support by Swiss National Science Foundation grant 31003A_46137, by an EpiphysX RTD grant from SystemsX.ch, as well as by the University Priority Research Program in Evolutionary Biology at the University of Zurich. JB is supported by a European Research Council's Advanced Grant.

References

Aharoni, A., Gaidukov, L., Khersonsky, O., McQ Gould, S., Roodveldt, C., & Tawfik, D. S. (2005). The evolvability of promiscuous protein functions. *Nat. Genet.*, 37:73-76.

Andreazzi, C. S., Thompson, J. N., & Guimarães, P. R. (2017). Network structure and selection asymmetry drive coevolution in species-rich antagonistic interactions. *Am. Nat.*,

190:99-115.

Barve, A. & Wagner, A. (2013). A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature*, 500:203-206.

Bascompte, J., Melián, C. J., & Sala, E. (2005). Interaction strength combinations and the overfishing of a marine food web. *Science*, 102:5443-5447.

Bascompte, J., Jordano, P., & Olesen, J. M. (2006). Asymmetric coevolutionary networks facilitate biodiversity maintenance. *Science*, 312:431-433.

Bersier, L.-F., Banasek-Richter, C., & Cattin, M. F. (2002). Quantitative descriptors of food-web matrices. *Ecology*, 83:23942407.

Dunne, J. A., Williams, R. J., & Martinez, N. D. (2002). Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecol. Lett.*, 5:558-567.

Fagan, W. F. & Hurd, L. E. (1994). Hatch density variation of a generalist arthropod predator: population consequenes and community impact. *Ecology*, 75:2022-2032.

Fenton, A., Antonovics, J., & Brockhurst, M. A. (2009). Inverse-gene-for-gene infection genetics and coevolutionary dynamics. *Am. Nat.*, 174:E230-42.

Fortuna, M. A., Zaman, L., Ofria, C., & Wagner, A. (2013). Evolving digital ecological networks. *PLoS Comput. Biol.*, 9:e1002928.

Fortuna, M. A., Zaman, L., Ofria, C., & Wagner, A. (2017). The genotype-phenotype map of an evolving digital organism. *PLoS Comput. Biol.* 13(2):e1005414.

Gould, S. J. & Vrba, E. S. (1982). Exaptation: a missing term in the science of form. *Paleobiology*, 8:4-15.

Guimarães, P. R., Jordano, P., & Thompson, J. N. (2011). Evolution and coevolution in mutualistic networks. *Ecol. Lett.*, 14:877-855.

Hosseini, S.-R., Martin, O. C., & Wagner, A. (2016). The potential for non-adaptive origins of evolutionary innovations in central carbon metabolism. *BMC Systems Biology*, 10:97.

Kokkoris, G. D., Troumbis, A. Y., & Lawton, J. H. (1999). Patterns of species interaction strength in assembled theoretical competition communities. *Ecol. Lett.*, 2:70-74.

Koonin, E. V. (2004). A non-adaptationist perspective of evolution of genomic complexity or the continued dethroning of man. *Cell Cycle*, 3(3):280-285.

LaBar T & Adami, C. (2016). Different evolutionary paths to complexity for small and large populations of digital organisms. *PLoS Comput. Biol.* 12(12):e1005066.

Lynch, M. & Conery, J. S. (2003). The origins of genome complexity. *Science*, 302:1401-1404.

McCann, K., Hastings, A. & Huxel, G. R. (1998). Weak trophic interactions and the balance of nature. *Nature*, 395:794-798.

Meyer, J. R., Dobias, D. T., Weitz, J. S, Barrick, J. E., Quick, R. T., & Lenski, R. E. (2012). Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science*, 335:428-432.

Neutel, A., Heesterbeek, J. A. P., & Ruiter, P. C. (2002). Stability in real food webs: weak links in long loops. *Science*, 296:1120-1123.

Notebaart, R. A., Szappanos, B., Kintses, B., Pálb, F., Györkeib, A., *et al.*. (2014). Network-level architecture and the evolutionary potential of underground metabolism. *Proc. Natl. Acad. Sci. USA*, 111:11762-11767.

Nuismer, S. L., Jordano, P., & Bascompte, J. (2013). Coevolution and the architecture of mutualistic networks. *Evolution*, 67:338-354.

O'Brien, P. J. & Herschlag, D. (1999). Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.*, 6:R91- R105.

Ofria, C. & Wilke, C. O. (2004). Avida: a software platform for research in computational evolutionary biology. *Artif. Life*, 10:191-229.

Paine, R. T. (1999). Food-web analysis through field measurement of per capita interac-

tion strength. *Nature*, 355:73-75.

Prigogine, I. (1980). *From Being To Becoming*. Freeman, San Francisco CA.

Raffaelli, D. & Hall, S. (1995). Assessing the relative importance of trophic links in food webs. In *Food webs: integration of patterns and dynamics*. Polis, G. & Winemiller, K. (eds). Chapman & Hall, New York, pp.185-191.

Ray, T. S. (1991). An approach to the synthesis of life. In: Langton, C., C. Taylor, J. D. Farmer, & S. Rasmussen (eds). *Artificial Life II*. Addison-Wesley, Redwood City, CA.

Wagner, A. (2017). The White Knight hypothesis, or does the environment limit evolutionary innovations? *Trends in Ecology and Evolution*, 32:131-140.

Wilke, C. O., Wang, J. L., Ofria, C, Lenski, R. E., & Adami, C. (2001). Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, 412:331-333.

Wilke, C. O. & Adami, C. (2002). The biology of digital organisms. *Trends. Ecol. Evol.*, 17:528-532.

Wootton, J. T. (1998). Estimates and tests of per capita interaction strength: diet, abundance, and impact of intertidally foraging birds. *Ecol. Monogr.*, 67:45-64.

Zaman, L., Devangam, S., & Ofria, C. (2011). Rapid host-parasite coevolution drives the production and maintenance of diversity in digital organisms. In *Proceedings of the* 13^{th}

Annual Conference on Genetic and Evolutionary Computation. Dublin, Ireland, pp.219-226.

Zaman, L., Meyer, J. R., Devangam, S., Bryson, D. M., Lenski, R. E., & Ofria, C. (2014). Coevolution drives the emergence of complex traits and promotes evolvability. *PLoS Biol.*, 12:e1002023.

Data accessibility.

Data are availaible in the Supporting Information.

Author contributions.

Conceived and designed the experiments: MAF, JB, and AW.

Performed the experiments: MAF.

Analyzed the data: MAF.

Contributed reagents/materials/analysis tools: LZ.

Wrote the manuscript: MAF, LZ, JB, and AW.

Competing interests

The authors declare no competiting interests.

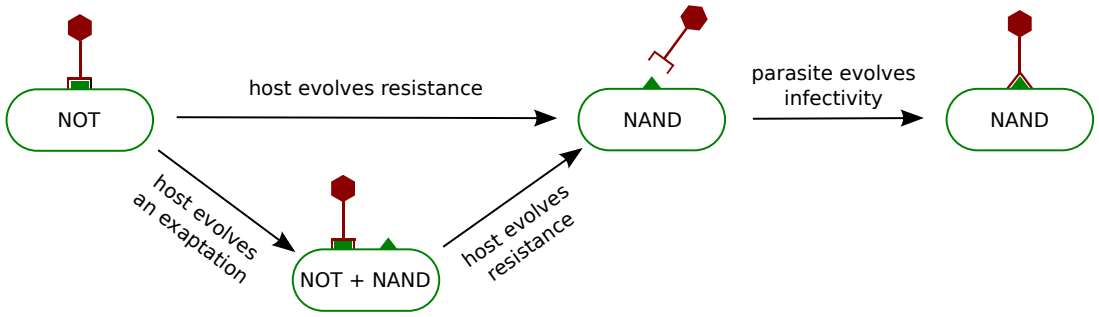


Figure 1: Schematic representation of the adaptive and non-adaptive origins of a host resistance trait. The host (green ellipse) has a receptor on its surface (i.e., a trait represented by a green rectangle) which allows it to consume resources from the environment. That receptor is also used by a parasite (in red) to infect the host. Computing logic functions, such as NOT and NAND, by executing the instructions that both digital hosts and parasites harbor in their genomes is analogous to having resistance and infectivity traits. The host can evolve a resistance trait (i.e., computing a novel Boolean logic operation such as NAND, represented by a green triangle on the surface of the host) following two different evolutionary trajectories: 1) evolving the novel trait through a single-point mutation while loosing the ancestral one (i.e., computing the NAND logic function while loosing the ability to perform the NOT logic function); or 2) evolving a non-adaptive trait (i.e., exaptation) through a single-point mutation that does not confer resistance to the host but facilitates the loss of the ancestral trait through a second mutation and, hence, provides resistance to the parasite. In this study, we compare a scenario with both possibilities (adaptive and non-adaptive) versus a second scenario where we suppress non-adaptive origins. The difference between the two scenarios will help us estimate the importantce of non-adaptive origins of evolutionary innovations. After either resistance was evolved, the parasite may overcome it by mutations in its own genome and become infective again.

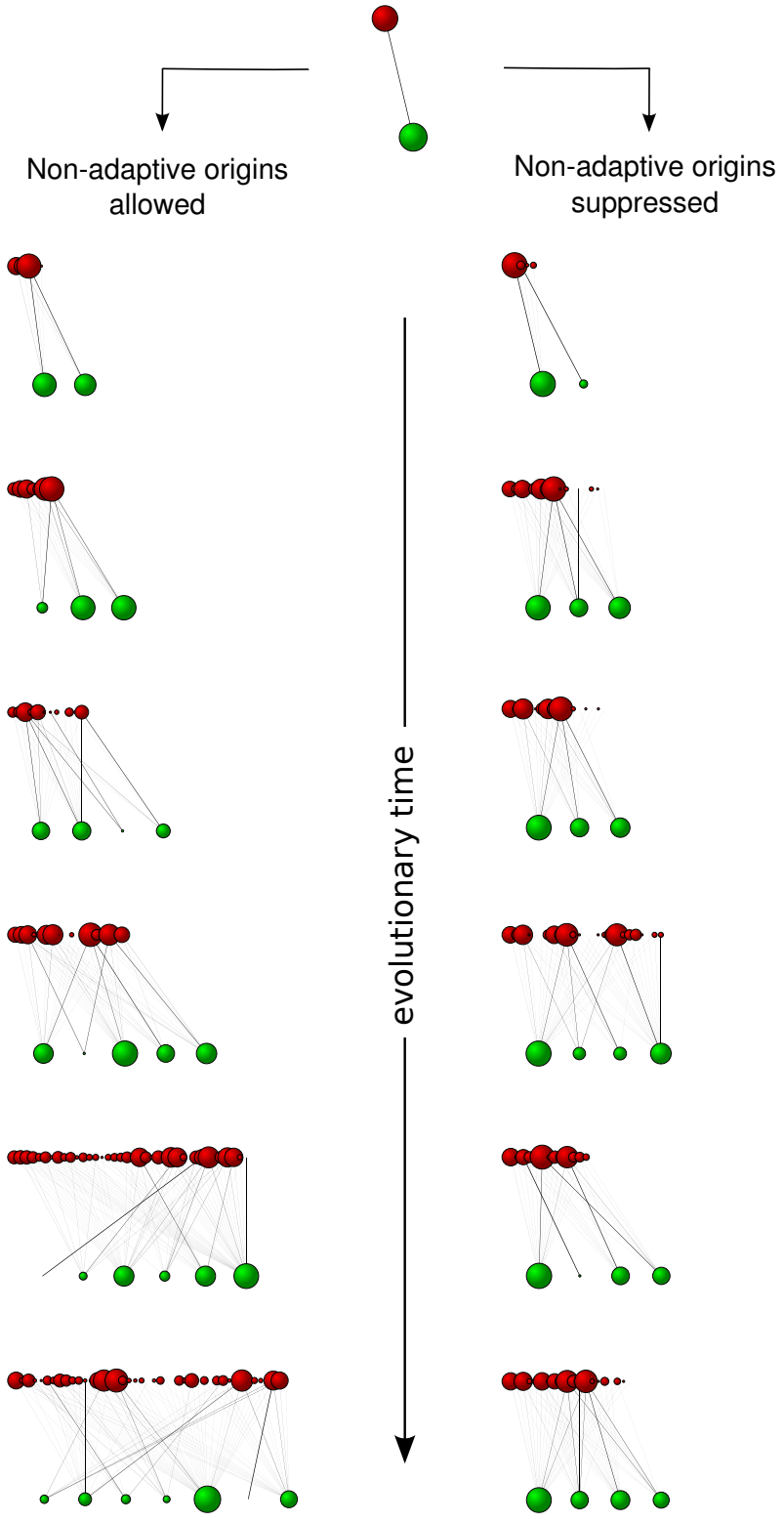

Non-adaptive origins allowed
Non-adaptive origins suppressed
evolutionary time

Figure 2: Evolving interaction networks among digital host and parasite phenotypes in two contrasting scenarios: (i) when resistance traits have both adaptive and non-adaptive origins (left), and (ii) when resistance traits have only adaptive origins (right). A phenotype (depicted as either a green or red node for hosts and parasites, respectively) is defined by a unique combination of Boolean logic functions (i.e., traits) that organisms with that phenotype compute. Starting from the same ancestral host and parasite phenotypes (represented at the top), the diversification of host resistance traits as well as who infects whom (depicted by the links connecting the nodes) are tracked over evolutionary time. The width of each link (i.e., interaction strength) is proportional to the fraction of organisms having a particular host phenotype infected by parasites encoding a given phenotype. The size of the nodes is proportional (in logarithmic scale) to the number of organisms having a particular phenotype. Each network depicts host-parasite interactions recorded at a specific point in evolutionary time (equal time for both right and left boxes placed at the same height). Only single-trait host phenotypes were allowed to evolve when traits had only adaptive origins. For traits with also non-adaptive origins, only single-trait hosts are represented. Parasite phenotypes can have more than one trait in both scenarios.

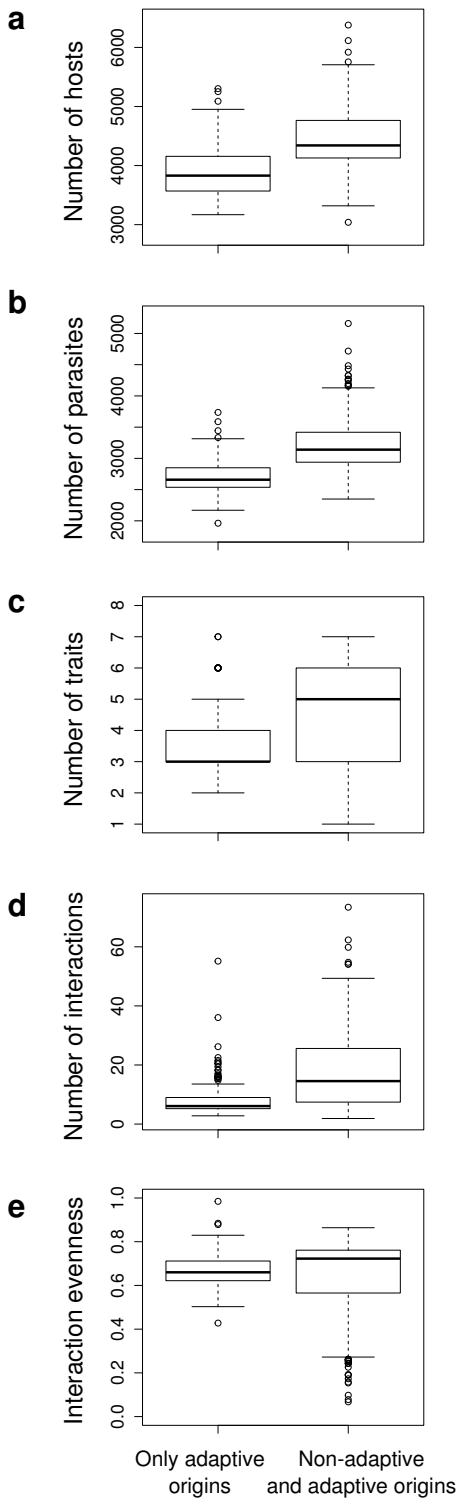

a
Number of hosts
3000
4000
5000
6000
b
Number of parasites
2000
3000
4000
5000
c
Number of traits
1
2
3
4
5
6
7
8
d
Number of interactions
0
20
40
60
e
Interaction evenness
0.0
0.2
0.4
0.6
0.8
1.0
Only adaptive origins
Non-adaptive and adaptive origins

Figure 3: Ecological and evolutionary consequences of the non-adaptive origins of host resistance traits. The number of hosts (a), parasites (b), traits in the host population (c) and interactions of single-trait hosts (d), were higher when resistance traits have both adaptive and non-adaptive origins than when the traits have only adaptive origins. On the contrary, interaction evenness (i.e., how well distributed the interaction strengths are among hosts and parasites accounting for differences in the number of interactions) was lower when resistance traits have both adaptive and non-adaptive origins than when the traits have only adaptive origins (e). Median (n=193 coevolutionary processes among host and parasite phenotypes) and upper and lower quartiles are shown.

Supporting Information

Sampling genotype space

We first generated 50 viable organisms by selecting random genomes with 100 instructions, where we chose each instruction in a genome randomly and uniformly among a set of 32 instructions out of the 33-instruction alphabet (Supplementary File 5). We excluded the "Inject" instruction from the instruction set at this stage because this instruction is only used by parasites to infect hosts, and hence, it does not play any role in the process of generating organisms that will become hosts. Finding those 50 organisms required us to sample $\approx 10^{11}$ random sequences. These 50 organisms were only capable of self-replication and did not perform any Boolean logic operation (i.e., they were merely viable organisms; Supplementary File 6).

In a second step, we aimed to find organisms capable of performing the NOT logic operation. To this end, for each of those merely viable organisms we performed 50 random walks through the genotype space. Each step in each random walk mutated one randomly chosen instruction in the random-walking genotype, and replaced it with a randomly-chosen instruction from the 32-instruction set mentioned above. Whenever such a mutation produced a non-viable organism or an organism capable of performing any logic function distinct from NOT, we reverted the mutation and mutated a new, randomly chosen instruction. We repeated this procedure until a viable organism capable of performing the NOT logic function appeared (Supplementary File 7).

Finally, to increase the robustness of those single-trait organisms, we performed a purifying selection process that moved organisms away from the merely-viable set of genotypes. That is, for each organism we started an evolutionary process in the standard mode of Avida from a population consisting of 10^4 organisms having the previously encountered

genotype. This process neither imposed selection for reducing the number of executed instructions required to produce an offspring (i.e., generation time) nor for performing any Boolean logic operations. Only organisms performing the NOT logic operation were allowed to replicate and their offspring differed in a single randomly chosen instruction from the genotype of its parent. After 10^5 updates, where an update is the amount of time during which an organism executes on average 30 instructions, we stopped the evolutionary process and kept only one randomly chosen genotype per population. By imposing this high rate of mutations, we evolved organisms with large robustness (Wilke *et al.*, 2001). In summary, our sampling of the genotype space found 50 mutationally-robust viable organisms capable of performing the NOT logic operation (Supplementary File 8).

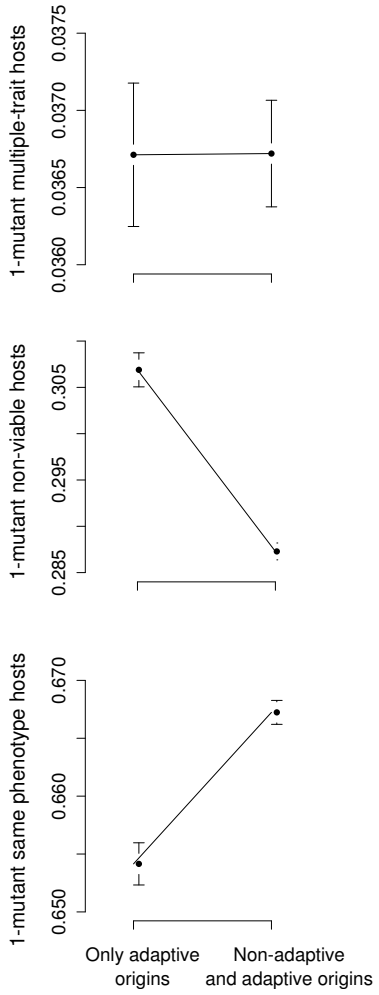
Searching for host and parasite ancestors

A parasite must replicate faster than its host, otherwise it will die when the host replicates. Then, we took the 25 organisms with the longest generation times as hosts. After that, we replaced the instruction "Divide-Erase" (i.e, the instruction that splits the organism's offspring from its parent and places it in a new memory space) by the instruction "Inject" (i.e., the instruction that the parasite uses to place its offspring into a host) from the genomes of the remaining 25 organisms. This is a simple way to convert an organism into a parasite. But the viability of a parasite depends on the host as well, since the parasite must replicate after the host has performed the logic operation NOT in order to be able to infect it. We tested the viability of those 25 parasites by observing their ability to replicate and infect each of the 25 hosts in the standar mode of Avida. Only 15 parasites were able to replicate and infect at least one of the 25 hosts (Supplementary File 9). None of the 5 parasites with the longest generation times were viable on any of those 25 hosts. We took those 5 parasites and replaced the instruction "Inject" by the instruction

"Divide-Erase" to revert them into hosts again, and added them to our previous set of hosts, which finally comprises a total number of 30 hosts (Supplementary File 10).

In the next step, we quantified the long-term stable coexistence of all $30 \times 15 = 450$ host-parasite pairs in a purely ecological scenario. That is, we performed 450 ecological processes in the standard mode of Avida where mutations were allowed neither in hosts, nor in parasites. Each process started from a population of 10^4 single-genotype hosts and, after the host reached its carrying capacity ($K \approx 6500$, on average for the 30 hosts, after 1000 updates) we infected half of the host population with the same ancestral parasite genotype. Then, we computed the number of host-parasite population pairs where both hosts and parasites coexisted after 10^4 updates. Only 166 (37%) host-parasite population pairs coexisted when the parasite virulence was set to 0.75. This number increased to 216 (48%) pairs when virulence was 0.90. We used those 216 host-parasite pairs as the ancestors of the evolutionary processes analyzed in this study (Supplementary File 2).

Supplementary Figures



Supplementary Figure 1: Analysis of the 1-mutant neighborhood of host populations. We computed the fraction of the $32 \times 100 = 3200$ 1-mutant neighbors that are i) non-viable, ii) viable and have the same phenotype than the parent organism, and iii) viable and have a novel phenotype consisting of more than one trait, for each host at different time steps (data pooled together) and for each scenario (i.e, when traits had both adaptive and non-adaptive origins, and when they had only adaptive origins).

Supplementary Files

S1 File. 33-instruction set.

S2 File. Host-parasite pairs used as ancestors.

S3 File. Configuration of the non-adaptive and adaptive scenario.

S4 File. Configuration of the adaptive scenario.

S5 File. 32-instruction set.

S6 File. Merely-viable organisms.

S7 File. Viable organisms capable of computing the NOT logic function.

S8 File. Viable and mutationally-robust organisms capable of computing the NOT logic function.

S9 File. Parasites.

S10 File. Hosts.